Projection-free first-order methods for nonsmooth optimization

Antonio SILVETI-FALLS

Keywords

Constrained optimization, convex optimization, stochastic optimization, machine learning, Frank-Wolfe, Conditional Gradient, data science.

Context

Projection-free first-order optimization methods, such as the Frank-Wolfe algorithm [4] and conditional gradient methods [7], have proven to be useful for many machine learning and data science problems [1] due to their ability to handle complex constraint sets without requiring possibly expensive projection operations. These methods rely on solving a linear minimization subproblem over the feasible domain at each iteration, making them attractive for large-scale optimization problems [3]. However, their analysis was traditionally relegated to smooth objective functions.

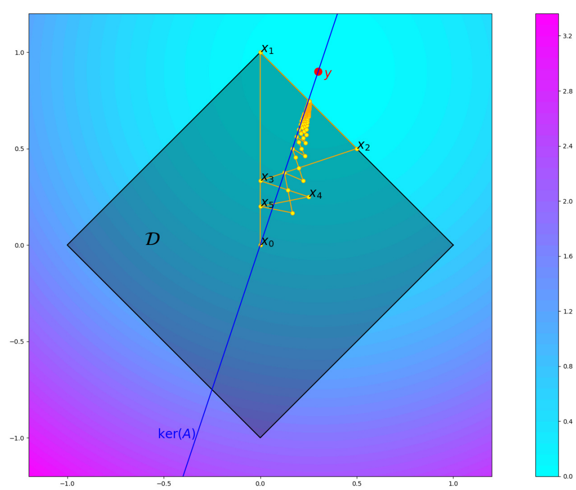

Figure 1: A nonsmooth variant of the Frank-Wolfe algorithm for solving the toy problem

$$\min_{x\in\mathbb{R}^2: Ax=0, \|x\|_1\le 1} \frac{1}{2}\|x-y\|^2$$

State of the Art

Recent works in this area have focused on minimizing objectives of the form $f+g$ over a convex, compact constraint set $\mathcal{C}$, where f is $C^{1,1}$ smooth (continuously differentiable with Lipschitz-continuous gradient) and convex, and g is nonsmooth but convex, proper, and lower semicontinuous [8, 9, 10, 11, 5]. This problem structure arises in various applications, including sparse recovery, matrix completion, and more. We have also come up with some preliminary results indicating that a Frank-Wolfe approach is also capable of tackling $f+g$ when f is possible nonconvex, which we plan to expand on in this project.

1 Novelty

Building upon our recent results, this post-doctoral research project aims to push the boundaries of nonsmooth projection-free optimization by exploring several innovative directions:

1. **Adaptive step sizes and smoothing**: We will investigate strategies to adapt the step size schedule based on the local geometry of the problem, utilizing a local curvature estimate. The goal is to accelerate convergence in practice. We also plan to investigate the smoothing schedule for the nonsmooth function g and its effect on convergence. Although convergence is guaranteed

under abstract summability assumptions on the parameters [8], the effects of different choices of initial values or sequences is not yet understood.

2. **Accelerated algorithms**: We will explore accelerated variants related to the Conditional Gradient Sliding [6] and Boosted Frank-Wolfe [2] approaches, which utilize more than one call to the linear minimization oracle per iteration. The aim will be to prove convergence rates and investigate their performance numerically.

2 Objectives

The primary objective of this post-doctoral research project is to find new and analyze new variants of projection-free optimization methods for nonsmooth problems. We aim to develop novel algorithms that improve upon the theoretical convergence rates and practical performance of existing techniques by leveraging adaptive step sizes, smoothing schedules, and multiple calls to the linear minimization oracle at each iteration.

Furthermore, we plan to demonstrate our theoretical claims in practice by implementing the proposed algorithms as efficient, open-source software packages in python. This will facilitate their adoption by the broader research community and enable their application to problems in various domains such as machine learning, signal processing, and imaging.

Desired profile

The desired candidate should have experience with theoretical analysis of optimization algorithms, in particular first-order methods. Experience and familiarity with vectorized Python programming, in particular with the common deep learning libraries (PyTorch, JAX, or TensorFlow) or at least with NumPy, will be necessary.

References

[1] Gábor Braun, Alejandro Carderera, Cyrille W Combettes, Hamed Hassani, Amin Karbasi, Aryan Mokhtari, and Sebastian Pokutta. Conditional gradient methods. *arXiv preprint arXiv:2211.14103*, 2022.

[2] Cyrille Combettes and Sebastian Pokutta. Boosting frank-wolfe by chasing gradients. In *International Conference on Machine Learning*, pages 2111–2121. PMLR, 2020.

[3] Cyrille W Combettes and Sebastian Pokutta. Complexity of linear minimization and projection on some sets. *Operations Research Letters*, 49(4):565–571, 2021.

[4] Marguerite Frank, Philip Wolfe, et al. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

[5] Gauthier Gidel, Fabian Pedregosa, and Simon Lacoste-Julien. Frank-wolfe splitting via augmented lagrangian method. In *International Conference on Artificial Intelligence and Statistics*, pages 1456–1465. PMLR, 2018.

[6] Guanghui Lan and Yi Zhou. Conditional gradient sliding for convex optimization. *SIAM Journal on Optimization*, 26(2):1379–1409, 2016.

[7] Evgeny S Levitin and Boris T Polyak. Constrained minimization methods. *USSR Computational mathematics and mathematical physics*, 6(5):1–50, 1966.

[8] Antonio Silveti-Falls, Cesare Molinari, and Jalal Fadili. Generalized conditional gradient with augmented lagrangian for composite minimization. *SIAM Journal on Optimization*, 30(4):2687–2725, 2020.

[9] Antonio Silveti-Falls, Cesare Molinari, and Jalal Fadili. Inexact and stochastic generalized conditional gradient with augmented lagrangian and proximal step. *Journal of Nonsmooth Analysis and Optimization*, 2(Original research articles), 2021.

[10] Alp Yurtsever, Olivier Fercoq, Francesco Locatello, and Volkan Cevher. A conditional gradient framework for composite convex minimization with applications to semidefinite programming. In *International Conference on Machine Learning*, pages 5727–5736. PMLR, 2018.

[11] Alp Yurtsever, Suvrit Sra, and Volkan Cevher. Conditional gradient methods via stochastic path-integrated differential estimator. In *International Conference on Machine Learning*, pages 7282–7291. PMLR, 2019.